DOCUMENT RESUME

ED 377 979 PS 022 925

AUTHOR Honig, Alice Sterling
TITLE Evaluation of Early Childhood Enrichment Programs.
PUB DATE Dec 94
NOTE 20p.; Paper presented at the Annual Conference of the National Association for the Education of Young Children (Atlanta, GA, November 30-December 3, 1994).
PUB TYPE Information Analyses (070) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Early Childhood Education; Educational Attitudes; Educational Quality; *Enrichment Activities; *Evaluation Criteria; *Evaluation Methods; Evaluation Needs; *Evaluation Problems; Program Attitudes; *Program Evaluation; Program Improvement

ABSTRACT

Leaders in the field of enrichment programs for young children and their families often have difficulty when considering the nature and utility of evaluations of such programs. Decisions as to how much evaluation to do, how to choose instruments, and how to train testers, observers, and interviewers become crucial for establishing not only the technical dimensions of evaluation but also the value of the resulting evaluation. Choice of assessment measures or procedures often depends on: (1) the program's need for formative evaluations; (2) the philosophy or theoretical orientation of the program; (3) the amount of financial support available for evaluation; and (4) the degree of data collection obtrusiveness permitted by staff. Some evaluations focus on detailing the quality of a program, while others focus on the children's learning environment or the children themselves. Timing is a crucial factor in evaluations, and long-term impact and family variables can also be included in program evaluations. Specific assessments should be used that relate to the purposes of the program and the specific intervention carried out. Overall, flexibility and creativity in choice of assessments can enrich the lives of children and help program staff and parents accommodate their needs. (Contains 19 references.) (MDM)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 377 979

EVALUATION OF EARLY CHILDHOOD ENRICHMENT PROGRAMS[1]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Alice Sterling Honig, Ph.D.

Syracuse University

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Alice Sterling Honig
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

Leaders in the field of enrichment programs for young children and their families often have a difficult time when considering the pros and cons of program evaluation. Their major energies are focused on thinking through programmatic philosophy and implementation procedures to enhance children's lives. Yet when evaluation components are carefully built into the planning process, they can often serve as a powerful adjunct to enhance the quality of service provision. Decisions as to how much evaluation to do, how to choose instruments, and how to train testers, observers, and interviewers become crucial for establishing not only the technical dimensions of evaluation but also the value of evaluation. When leaders are clear and convincing, then staff realizes how important this component will be in helping a program meet its targeted goals for families; they may change from suspicion of evaluation to enthusiastic support.

Program evaluations have many dimensions. Systematic efforts to evaluate an enrichment or intervention program requires much decision making. Depending on the **goals** of the evaluation, the form and focus, intensity and extensiveness of the procedures and the

[1]. Paper presented in a forum entitled: " Leadership in early childhood education from the 1960s to the 1990s and beyond" at the annual conference of the National Association for the Education of Young Children, Atlanta, Georgia, December, 1994.

PS 022925

BEST COPY AVAILABLE

level of formality will vary (Honig, 1995).

Choice of Evaluation Measures

Choice of assessment measures or procedures often depends on the following factors:

1. Program's need for formative evaluations in order to provide frequent periodic feedback useful for further staff training. This must be balanced by the necessity of collecting summative evaluation in order to convince funding agencies of the ultimate efficacy of the program.

2.Philosophy or theoretical orientation of the director and the evaluator. Decisions will have to be made about mixing quantitative and qualitative evaluations. How much time can staff spend, for example on writing in-depth running records of child interactions in contrast with the usefulness of more time-limited checklists for assessing child activities and interactions?

2.Financial support available for more elaborate or more modest evaluations. If a program has practically no resources for evaluation, the Director can approach a local college for collaboration. Then students in child development and early education courses can carry out assessments. Students will profit from hands-on experiences of observing and assessing. Their findings will serve the program well with valuable data and insights. They can suggest measures of child development, whether in language, classroom learning, or positive social interpersonal skills.

Some classroom curricula come with built in assessment tools.

Shure (1993) has created programs called "I Can Problem Solve" for teachers of young children. She also provides assessment scenarios such as the WHNG (What Happens Next game)to use with each child. Armed with the children's responses, a teacher can decide where each child is in terms of positive problem solving skills.

3.The degree of data collection obtrusiveness permitted by staff, Boards of Directors, and parents. Outside evaluators may choose to carry out in-depth interviews or stage stressful problem solving situations to use with parents. A director serving teen parents in a program providing infant/toddler childcare plus classes for young mothers could feel anxious that overly intrusive inquiries will cause some of the teen parents to drop out of the program. That director may opt for using naturalistic observations.

Unobtrusive observations can be quite effective. In one program where I advised staff, a teen mother would growl " Shut up you" as she changed her infant when ready to take him home, after the childcare program and her QED schooling in the same building were over for the day. As programmatic efforts to provide supports and insights for the young mothers continued, and teacher modeling of gentle, empathic care was observed daily by the mothers, then changes were observed in maternal diapering table behavior at the end of the day.

4.Programs needs differ in choice of keeping longitudinal, ongoing records or briefly sampling children's behaviors from time to time. Sometimes programs must make tradeoffs. They may trade off losing in-depth long-term outcome measures on a few children for a more

comprehensive collection of outcomes with greater coverage of more children.

Where financial resources permit, an evaluation team may decide to enhance external validity, which means that findings can be generalized to larger groups, such as children of different ethnicities and family status. Large scale projects collect data on a national stratified sample that is geographically and ethnically representative as in the study of childcare staffing patterns and quality of care (Whitebook et al, 1989). There, stability and quality of care were found significantly tied to staff salary and child development training. Yet time and budget constraints may permit only a one-time collection of specific events, such as children's separation anxiety distress and how each episode is resolved. Or staff can carry out brief time samplings of behaviors of importance with respect to project goals.

Focus on the Quality of the Program Environment

Some evaluations focus on specifying in detail the quality of program. They make explicit the organization and structural components of a program. In residential nurseries in England, hierarchical characteristics of the institution impacted on child language outcome measures. The more rigidly caregivers were dependent on a director's decisions, the less competent the children were on the Reynell scales of receptive and expressive language. When children were cared for by caregivers who were given more autonomy and flexibility in deciding their own daily schedules, then the children's Reynell language scores were higher.

Unfortunately, even though the evaluation focus was on the effects of institutional organization on children's language, analysis showed that when hierarchy was more rigid, then staff turnover and instability was also greater. Often, evaluation teams will find that hidden variables impact on the outcomes chosen for measurement and need to be considered and taken into account in advance by evaluators (Tizard et al., 1972) .

Many evaluations of program set as first priority the measurement of the environment for children. That is, on paper fancy program goals and professed adherence to developmentally appropriate practices (Bredekamp, 1987) may look impressive. Formative evaluations inquire whether the program activities and interactions actually match the stated goals. Attention to such evaluations may prove significant over time. Primary school children in Trinidad who had attended a more definitively teacher-directed rather than child-centered preschool program had lower mean achievement scores and were less likely to tell important events to their teachers and to concentrate in class (Kutnick, 1995).

The way in which learning areas are structured, the movement of children from one activity to another, the amount of teacher dominated vs. child choice that is reflected in the ongoing daily activities may all be the focus of evaluation. Some instruments, such as the ECRS developed by Harms & Clifford (1980) focus on the adequacy of the care setting including classroom furnishings, personal care routines, and creative activities.

For the Family Development Research program, in Syracuse, New York, we developed an observation checklist technique, ABC (Adult Behaviors in Caregiving) to assess caregiver functioning. Every two minutes a teacher is checked for whatever behaviors she is carrying out with a boy or girl in different curricular domains, such as Piagetian tasks, promoting prosocial interactions, reading, and soothing. The three easy-to-use ABC checklists (for teachers of infants, toddlers and preschoolers) proved to be sensitive indicators of the efficacy of an intensive inservice teacher training week held every autumn (Honig & Lally, 1988).

Focus on the Target Child

Many evaluations focus on changes in the children served. Measures include achievement tests, on-task performance rates, positive or inappropriate socioemotional interactions and behaviors (with peers and with adults), and cognitive competencies often defined via IQ or developmental scores on psychometric tests. In recent years there have been vigorous efforts to change from product-oriented evaluations of test results to process-oriented evaluations of the ongoing work of the child, such as drawings and dictated stories.Genishi (1992) urges that such assessments are more naturally and conceptually linked to curriculum..

When the focus of the evaluation is on the outcomes of program for individual children, then decisions must be made whether to assess one or more particular domains of functioning, such as cognition. For example, in the early evaluations of Head Start, minor intellectual gains were found that washed out by third grade.

Yet the percent of children whose medical problems were identified and remediated during the Head Start years was impressive as a measure of success. Long-term results show that graduates were less likely to have a history of delinquency or a criminal record, and for girls, less likely to be teenage mothers (Schorr, 1988).

The focus of evaluation can be on the child's interpersonal relationships, whether child-teacher or child-peer relationships, and the ability to solve social spats with positive resolutions. In Title XX schools, highly successful low-income kindergarten children had more harmonious relationships with family, carried out required chores, were read to regularly at home, and had a father in the home. Their teachers reported that they were hard working, articulate, with a sense of humor, persistently on task in the classroom, and able to solve their social altercations peaceably with peers (Swan & Stavros,1973).

The Importance of Time and Length of Data Collection

Timing is important in evaluation. In a study of Chicago low SES preschoolers, examiners tested children immediately upon entry to the program. Later they found impressive increases in scores at the end of the school year. The next year, examiners waited to do initial testing until the new group of preschoolers were thoroughly comfortable in the school setting. Not surprisingly, the initial scores of the second group of preschoolers were much higher than initial scores for the first group, and the year-end effects of their program participation did not look as impressive.

Some evaluators want to know how children are faring right

after a program ends. Others are far more concerned with long term effects. How well will the children's newly acquired skills or higher IQ scores hold up years after the program ends? Some program effects wash out early. Others, such as giving a child a positive motivation for learning and a concept of teachers as loving, trustworthy, and helpful adults may result in more positive child school attitudes and class cooperation many years later (Lally et al., 1988).

Some evaluations that are concerned with longer term effects of program on family functioning, will focus on the younger siblings of target children. Such vertical diffusion effects were found by Dr. Susan Grey in her DARCEE project. The Milwaukee Project for infants and preschoolers of intensely at-risk families living in dilapidated housing also showed consistently higher IQ scores for younger siblings of the target children as well as impressive IQ gains for target children immediately post program. However, in that project, despite costly early intervention with the children from early infancy, the youths at end of high school were defiant, truant, and doing poorly. They had attended high poverty inner city schools from the time they left the intervention program and began public school. Evaluations that test for effects immediately post intervention may miss both positive sleeper effects and disappointing washout effects. Long term evaluations may give confidence that a program's gains will be sustained years after the child has graduated.

Family Variables: Hidden Impact on Preschool Program Outcomes

Sometimes evaluators find that there appears to be no difference between children in experimental (enriched) preschool programs and their controls. When family variables are taken into account, however, then the positive effects of program become clearer. Levenstein (1988) describes the difference between Hesitator and Striver mothers in her Home Visitation project. Both groups of mothers had babies and dropped out of school. But Striver mothers subsequently went back for GED diplomas and enrolled in work or study programs for themselves. On a long term basis, their children were not significantly impacted by the Mother-Child Home Visitation two-year program that weekly brought books and toys to the toddlers' homes. However, children of the Hesitator mothers (who had not galvanized themselves toward either self or family improvement) did significantly better than control youngsters even many years after their participation in the MCHV program. In the United Kingdom, Meadow & Cashdan (1988) similarly report that the most socially disadvantaged children who received preschool education benefitted the most, and these differences were reported to last when children were assessed at age 10.

Control Groups

Many evaluators of program do not have funds for a research design that includes a control group. Post-program outcomes are assessed without regard for possible effects of increased child maturity, or improvement in family functioning, or other hidden variables not directly related to programmatic inputs (Honig, 1983).

When children are randomly assigned to a control group prior to carrying out the enrichment program, then a more powerful test can be made of the hypothesis that the program made a difference in child outcomes.

Often, random assignation of children to program or control groups is not possible. It may be ethically or politically unwise to refuse some youngsters access to a high quality preschool program. Later on, careful matching of a group of control subjects with experimental children then becomes necessary. Age, sex, ethnicity, birth order, income, marital arrangement, number and spacing of siblings, are important variables to take into account in careful matching.

If longitudinal data are to be gathered, one pitfall of using control groups, whether they are matched carefully or chosen earlier through random assignation is that families who were assigned to the control group may be differentially lost through attrition. Then comparisons become exceedingly difficult, as only the most cooperative and highly motivated control families are being compared with program graduates. By sending birthday cards and holiday greeting cards and by periodic friendly telephone follow-ups, programs may maintain contact with families and prevent the attrition problem that often plagues evaluators in longitudinal follow-up studies.

Specific vs. General Program Effects

If an enrichment program focuses on language enhancement, then use of a general IQ test would be inappropriate. Specific

assessments should be used that relate to the purposes of the program and the specific interventions carried out. Suppose a language enrichment program schedules pre and post program assessments. Suppose also that the children are bilingual and have been freely allowed to use their native languages without anyone correcting their English grammar. Then using all the subtests of the ITPA (Illinois Test of Psycholinguistic Ability), including the grammatic closure subtest as an evaluation measure, would be inappropriate. Measures to reflect change should be related to the actual curricular efforts that are undertaken.

Screening vs. Psychometric Measures

Many enrichment programs are a first line of defense against the risk of school failure later on for children from at-risk families. Staff may not be as concerned about how high children's IQs are, but they are concerned lest any of the children need particular targeted specialist services. Concerned caregivers can begin with an easy-to-learn screening tool, such as the Denver Developmental Screening Test. If the child fails two or more items in two or more of the four areas tested, then further more refined assessments may be needed. Staff will want to learn to use screening tests themselves, even when they may have to call in specialists if initial screening confirms that a child needs more specialized help.

Evaluations When Program Focus is Primarily on Parents

Some evaluation efforts will focus primarily on parents, because a major program goal, as in Home Start, is to empower the

parent. Home visitors provide insights, personal supports, information, books, developmentally appropriate toys, and social skills (such as positive discipline techniques) that will help parents parent more effectively in at-risk families (Honig, 1979). Vary evaluation techniques creatively in such cases. The fact that a parent now knows how to reach out and find appropriate social services or is using a library regularly to find books to read with the child may be excellent outcome measures of the success of program (Honig, 1979). The current presence of a stable and positive fathering figure for a young child can be a positive measure of the effectiveness of the program's impact on family.

Standardized tests and measures are not the only way for programs to reveal their positive accomplishments. In the Family Development Research Program (FDRP) in Syracuse, we counted our work as achieving positive changes when a mother was able to respond positively to our expressed admiration for the child during a home visit or was eating meals and talking with the child more frequently without the TV on at dinner time. Such items may be useful for assessing how well a family-focused program is meeting its goals. Items from the IPLET (Implicit Parental Learning Theory interview) and WHVR (Weekly Home Visit Report) measures from the Syracuse FDRP program as well as Dr. Bettye Caldwell's HOME Inventory can be helpful for evaluators searching for innovative measures of positive change in family functioning.

Who Are Your Data Gatherers?

When psychometric tests such as the Stanford Binet are chosen

as evaluation measures, then the testers must be thoroughly trained and capable clinicians. They should be caring and intimate in relating to young children and deeply appreciative of a child's cooperation. Hastily trained or ill-prepared testers without knowledge of how to interact effectively with young children cannot be trusted to gather reliable and valid data despite the fact that they may have "learned" the rudiments of the items to be presented in a given battery of tests.

Optimal testing is fervently to be desired. That means every child is well rested and well fed before being tested. This may sometimes mean breaking up a long series of tests over several days, or feeding a youngster, or taking a break with toys in a playroom, or even taking a walk around the block in the fresh air before continuing a battery of tests (Honig & Lally, 1989).

When paraprofessionals are trained to collect formative evaluation data after weekly home visits, frequent meetings may be necessary in order to make sure that no drift in operational definitions of observed or inquired items has occurred.

When classroom caregivers are required to carry out assessments in addition to their teaching, nurturing, program planning, and parent involvement efforts, then in-service training will be necessary. As far as possible, the instruments chosen or created should not put undue burdens on staff so that burnout does not occur.

Formative Evaluation As An Intervention

An important aspect of regular data gathering is that

caregivers and staff become intimately involved in and responsive to whether or not programmatic innovations are actually producing desired effects. True, some programmatic innovations take time to implement and time before effects are seen. When data collection becomes an integral aspect of program, then teachers have a stake in it. Outside evaluators may be seen as THEM vs. US. When caregivers themselves are observing, recording, pondering the meaning of child lack of responsiveness or becoming excited by child advances after worrisome delays, then evaluation becomes owned by the teachers. They have a stake in seeing that their work makes a positive difference in young children's lives. An additional advantage is that when screening or assessments are done in an ongoing fashion, even children from isolated and poorly socialized environments become accustomed to the rules and procedures of "testing" that staff carries out in loving and affirmative interactions with them.

When parents are invited to be present for assessments, the power of assessment as a further enrichment tool can be marked. Parents sharpen their observation skills. They begin to value children's tries instead of just "perfect" or "correct" scores. Parents can learn to model the genuine delight that a seasoned tester shows as she or he lures each young child into struggling with difficult problems and tasks on the cutting edge of learning. The parent learns how the Vygotskian "zone of proximal development" really works as the examiner assists the child in focusing on a task and supports a child's longer attention span and persistence

at the task. The importance of the adult as playmate and teaching companion in the child's learning process is clearly modelled as a parent observes a skilled examiner assist the child to perform optimally. A seasoned examiner even when working with a child who lags developmentally, will provide items that allow for clear cut child success as she or he attempts to assess basal and ceiling scores for a psychometric test. The examiner rejoices verbally and with clapping gestures at the young child' competent behaviors. A facilitative testing style is a highly developed skill. Teachers who learn to use achievement assessments, such as the Caldwell Preschool Inventory, will have the satisfaction of becoming acutely knowledgeable about just what domains the child has mastered, and where the child needs more sustained and helpful adult work toward new adventures in learning.

CONCLUSIONS

Flexibility and creativity in choice of assessments can enrich the lives of children rather than cause "test anxiety" to become entrenched early in a child's life. Child portfolios can be systematically gathered in ongoing evaluations, and evaluators can use both brief screening and more fine-tuned psychometric assessments judiciously. Well chosen evaluations help program personnel more clearly to decide where their efforts need boosting and where their strengths are evident in working with children and families.

Parents who are invited to sit in on assessment sessions where warm, intimate interactions take place between adult and child will

find rich rewards in getting to know their own children better and getting to appreciate small but significant advances in their young child's learning. Participating teachers will feel that they are "on top" of each individual child's learning patterns and abilities, so that they can uniquely individualize their program goals for each child.

REFERENCES

Bredekamp, S. (1987). (Ed.). Developmentally appropriate practice in early childhood programs serving children from birth through age 8. Washington, DC: National Association for the Education of Young Children.

Genishi, C. (1992). Ways of assessing children and curriculum: Stories of early childhood practice. New York, NY: Teachers College Press.

Harms, T., & Clifford, R. M> (1980). Early Childhood Environment Scale (ECRS). New York, NY: Teachers College Press.

Honig, A. S. (1979). Parent involvement in early childhood education. Washington, DC: National Association for the Education of Young Children.

Honig, A. S. (1983). Evaluation of infant/toddler intervention programs. In B. Spodek (Eds.), Studies in educational evaluation Vol.8 (pp. 305-316. London: Pergamon Press.

Honig, A. S. (1989). Longitudinal effects of quality preschool programs. Day Care and Early Education, 17,35-38.

Honig, A. S. (1995). Assessing the preparation of infant-toddler caregivers. In S. Reifel (Ed.) Advances in Early Education and Day Care.Vol.6. Greenwich, CT: JAI Press.

Honig, A. S., & Lally, J. R. (1981). Infant caregiving: A design for training. Syracuse, NY: Syracuse University Press.

Honig, A. S., & Lally, J. R. (1988). Behavior profiles of experienced teachers of infants and toddlers. In A. S. Honig

(Ed.), Optimizing early child care and education (Special Issue). Early Child Development and Care, 33, 181-199.

Honig, A. S., & Lally, J. R. (1989). Effects of testing style on language scores of four-year-old low-income "control" children in an intervention project. Early Child Development and Care, 41, 195-211.

Kutnick,P. (1994). Does preschool curriculum make a difference in primary school performance: Insights into the variety of preschool activities and their effects on school achievement and behaviour in the Caribbean Island of Trinidad: Cross sectional and longitudinal evidence. Early Child Development and Care, 103, 27-42.

Lally, J.R., Mangione, P., & Honig, A. S. (1988). The Syracuse University Family Development Research Program: Long-range impact of an early intervention with low income children and their families. In D. Powell (Ed.), Parent involvement as early childhood intervention: Emerging directions in theory, research, and practice (pp. 79-104). Norwood, NJ: Ablex.

Levenstein,P. (1988). The Mother-Child Home Program and the prevention of school disadvantage. Columbus, OH: Ohio State University Press.

Meadows, S., & Cashdan, A. (1988). Helping children learn. London: David Fulton .

Schorr, L. (1988). Within our reach: Breaking the cycle of disadvantage. New York: Anchor Doubleday.

Shure, M. (1992). I can problem solve: An interpersonal cognitive

problem-solving program (Vol.1 - preschool; Vol.2 - kindergarten and primary grades; Vol.3 elementary grades). Champaign, IL: Research Press.

Swan , R. W., & Stavros, H. (1973). Child rearing practices associated with the development of cognitive skills of children in low socioeconomic areas. Early Child Development and Care, 2, 23-38.

Tizard , B., Cooperman, O., Joseph, A., & Tizard, J. (1972). Environmental effects on language development: A study of young children in long-stay residential nurseries. Child Development, 43 , 337-358.

Whitebook, M., Howes, C., & Phillips, D. (1989). Who cares: Child care and the quality of care in America. Final report of the National Child Care Staffing Study. Oakland, CA: Child Care Employee Project.